

Bernoulli **16**(2), 2010, 561–584
DOI: [10.3150/09-BEJ209](https://doi.org/10.3150/09-BEJ209)

Uniform error bounds for a continuous approximation of non-negative random variables

CARMEN SANGÜESA

Departamento de Métodos Estadísticos, Facultad de Ciencias, Universidad de Zaragoza, Pedro Cerbuna, 12, 50009 Zaragoza, Spain. E-mail: csangues@unizar.es

In this work, we deal with approximations for distribution functions of non-negative random variables. More specifically, we construct continuous approximants using an acceleration technique over a well-know inversion formula for Laplace transforms. We give uniform error bounds using a representation of these approximations in terms of gamma-type operators. We apply our results to certain mixtures of Erlang distributions which contain the class of continuous phase-type distributions.

Keywords: gamma distribution; Laplace transform; phase-type distribution; uniform distance

1. Introduction

Frequent operations in probability such as convolution or random summation of random variables produce probability distributions which are difficult to evaluate in an explicit way. In these cases, one needs to use numerical evaluation methods. For instance, one can use numerical inversion of the Laplace or Fourier transform of the distribution at hand (see [2] for the general use of Laplace–Stieltjes transforms in applied probability or [9, 11] for the method of Fast Fourier Transform in the context of risk theory). Another approach is the use of recursive evaluation methods, of special interest for random sums (see [11, 18], for instance). Some of the methods mentioned above require a previous discretization step to be applied to the initial random variables when these are continuous. The usual way to do so is by means of rounding methods. However, it is not always possible to evaluate the distribution of the rounded random variable in an explicit way and it is not always clear when using these methods how the rounding error propagates when one takes successive convolutions. In these cases, it seems worthwhile to consider alternative discretization methods. For instance, when dealing with non-negative random variables, the following method ([10], page 233) has been proposed in the literature. Let X be a

This is an electronic reprint of the original article published by the ISI/BS in *Bernoulli*, 2010, Vol. 16, No. 2, 561–584. This reprint differs from the original in pagination and typographic detail.

random variable taking values in $[0, \infty)$ with distribution function F . Denote by $\phi_X(\cdot)$ the Laplace–Stieltjes (LS) transform of X , that is,

$$\phi_X(t) := Ee^{-tX} = \int_{[0, \infty)} e^{-tu} dF(u), \quad t > 0.$$

For each $t > 0$, we define a random variable $X^{\bullet t}$ taking values on $k/t, k \in \mathbb{N}$, and such that

$$P(X^{\bullet t} = k/t) = \frac{(-t)^k}{k!} \phi_X^{(k)}(t), \quad k \in \mathbb{N}, \quad (1)$$

where $\phi_X^{(k)}$ denotes the k th derivative ($\phi_X^{(0)} \equiv \phi_X$).

Thus, if we denote by L_t^*F the distribution function of $X^{\bullet t}$, we have

$$L_t^*F(x) := P(X^{\bullet t} \leq x) = \sum_{k=0}^{[tx]} \frac{(-t)^k}{k!} \phi_X^{(k)}(t), \quad x \geq 0, \quad (2)$$

where $[x]$ indicates the largest integer less than or equal to x . The use of this method allows one to obtain the probability mass function in an explicit way in some situations in which rounding methods could not (see, for instance, [4] for gamma distributions). Moreover, this method allows for an easy representation of L_t^*F in terms of F , which makes possible the study of rates of convergence in the approximation ([4, 5]). In [4], the problem was studied in a general setting, whereas in [5], a detailed analysis was carried out for the case of gamma distributions, that is, distributions whose density function is given by

$$f_{a,p}(x) := \frac{a^p x^{p-1} e^{-ax}}{\Gamma(p)}, \quad x > 0. \quad (3)$$

Also, in [16], error bounds for random sums of mixtures of gamma distributions were obtained, uniformly controlled on the parameters of the random summation index. In all of these papers, the measure of distance considered was the Kolmogorov (or sup-norm) distance. More specifically, for a given real-valued function f defined on $[0, \infty)$, we denote by $\|f\|$ the sup-norm, that is,

$$\|f\| := \sup_{x \geq 0} |f(x)|.$$

It was shown in [5] that for gamma distributions with shape parameter $p \geq 1$, we have that $\|L_t^*F - F\|$ is of order $1/t$, the length of the discretization interval. Note that $\|L_t^*F - F\|$ is the Kolmogorov distance between X and $X^{\bullet t}$, as both are non-negative random variables.

The aim of this paper is twofold. First, we will consider a continuous modification of (2) and give conditions under which this continuous modification has rate of convergence of $1/t^2$ instead of $1/t$ (see Sections 2 and 3). In Section 4, we will consider the case of gamma distributions to see that the error bounds are also uniform on the shape

parameter. Finally, in Section 5, we will consider the application of the results in Section 4 to the class of mixtures of Erlang distributions, recently studied in [19]. This class contains many of the distributions used in applied probability (in particular, phase-type distributions) and is closed under important operations such as mixtures, convolution and compounding.

2. The approximation procedure

The representation of L_t^*F in (2) in terms of a Gamma process (see [4]) will play an important role in our proofs. We recall this representation. Let $(S(u), u \geq 0)$ be a gamma process, in which $S(0) = 0$ and such that for $u > 0$, each $S(u)$ has a gamma density with parameters $a = 1$ and $p = u$, as given in (3). Let g be a function defined on $[0, \infty)$. We consider the gamma-type operator L_t given by

$$L_t g(x) := Eg\left(\frac{S(tx)}{t}\right), \quad x \geq 0, t > 0, \quad (4)$$

provided that this operator is well defined, that is, $L_t|g|(x) < \infty, x \geq 0, t > 0$. Then, for F continuous on $(0, \infty)$, L_t^*F in (2) can be written as (see [4], page 228)

$$L_t^*F(x) = L_t F\left(\frac{[tx] + 1}{t}\right) = EF\left(\frac{S([tx] + 1)}{t}\right), \quad x \geq 0, t > 0. \quad (5)$$

It can be seen that the rates of convergence of $L_t g$ to g are, at most, of order $1/t$ (see (40) below). Our aim now is to get faster rates of convergence. To this end, we will consider the following operator, built using a classical acceleration technique (*Richardson's extrapolation* – see, for instance, [9, 11]):

$$L_t^{[2]}g(x) := 2L_{2t}g(x) - L_tg(x) = 2Eg\left(\frac{S(2tx)}{2t}\right) - Eg\left(\frac{S(tx)}{t}\right), \quad x \geq 0. \quad (6)$$

We will obtain a rate of uniform convergence from $L_t^{[2]}g$ to g , of order $1/t^2$, on the following class of functions:

$$\mathcal{D} := \{g \in C^4([0, \infty)): \|x^2 g^{iv}(x)\| < \infty\}. \quad (7)$$

The problem with $L_t^{[2]}g$ is that when tx is not a natural number, $L_tg(x)$ is given in terms of Weyl fractional derivatives of the Laplace transform (see [6], page 92) and, in general, we are not able to compute them in an explicit way. However, if we modify $L_t^{[2]}g$ using linear interpolation, that is,

$$M_t^{[2]}g(x) := (tx - [tx])\left(L_t^{[2]}g\left(\frac{[tx] + 1}{t}\right)\right) + ([tx] + 1 - tx)\left(L_t^{[2]}g\left(\frac{[tx]}{t}\right)\right), \quad (8)$$

then we observe that the order of convergence of $M_t^{[2]}g$ to g is also $1/t^2$ on the following class of functions:

$$\mathcal{D}_1 := \{g \in C^4([0, \infty)): \|g''(x)\| \leq \infty \text{ and } \|x^2 g^{iv}(x)\| < \infty\}. \quad (9)$$

Moreover, the advantage of using $M_t^{[2]}g$ instead of $L_t^{[2]}g$ to approximate g is the computability. In the following result, we note that the last approximation applied to a distribution function F is related to L_t^*F , as defined in (2). From now on, \mathbb{N}^* will denote the set $\mathbb{N} \setminus \{0\}$.

Proposition 2.1. *Let X be a non-negative random variable with Laplace transform ϕ_X . Let $L_t^*F, t > 0$, be as defined in (2) and let $M_t^{[2]}F$ be as defined in (8). We have*

$$M_t^{[2]}F\left(\frac{k}{t}\right) = \begin{cases} F(0), & \text{if } k = 0, \\ 2L_{2t}^*F\left(\frac{2k-1}{2t}\right) - L_t^*F\left(\frac{k-1}{t}\right), & \text{if } k \in \mathbb{N}^*, \end{cases} \quad (10)$$

and

$$M_t^{[2]}F(x) = (tx - [tx])M_t^{[2]}F\left(\frac{[tx] + 1}{t}\right) + ([tx] + 1 - tx)M_t^{[2]}F\left(\frac{[tx]}{t}\right). \quad (11)$$

Proof. Let $t > 0$ be fixed. First, observe that by (8), we can write

$$M_t^{[2]}F\left(\frac{k}{t}\right) = L_t^{[2]}F\left(\frac{k}{t}\right), \quad k \in \mathbb{N}. \quad (12)$$

Now, using (6) and (4), we have $M_t^{[2]}F(0) = L_t^{[2]}F(0) = F(0)$, which shows (10) for $k = 0$. Finally, using (6), (4) and (5), we have, for $k \in \mathbb{N}^*$,

$$L_t^{[2]}F\left(\frac{k}{t}\right) = 2EF\left(\frac{S(2k)}{2t}\right) - EF\left(\frac{S(k)}{t}\right) = 2L_{2t}^*F\left(\frac{2k-1}{2t}\right) - L_t^*F\left(\frac{k-1}{t}\right). \quad (13)$$

Thus, (12) and (13) show (10) for $k \in \mathbb{N}^*$. Note that (11) is obvious by (8) and (12). This completes the proof of Proposition 2.1. \square

In the following example, we illustrate the use of the previous approximant in the context of random sums, defined in the following way. Let $(X_i)_{i \in \mathbb{N}^*}$ be a sequence of independent, identically distributed non-negative random variables. Let M be a random variable concentrated on the non-negative integers, independent of $(X_i)_{i \in \mathbb{N}^*}$. Consider the random variable

$$\sum_{i=1}^M X_i, \quad (14)$$

with the convention that the empty sum is 0.

Example 2.1. As pointed out in the [Introduction](#), an explicit expression for the distribution of (14) is usually not possible. Our aim is to consider an example in which this distribution can be evaluated explicitly and to compare our approximation method with some others considered in the literature. To this end, we consider that M follows a geometric distribution of parameter p , that is, $P(M = k) = (1 - p)^k p, k \in \mathbb{N}$ and $(X_i)_{i \in \mathbb{N}^*}$ are exponentially distributed (with mean 1, for the sake of simplicity). In this case, it is well known (use LS transforms, for instance) that (14) has the same distribution as a mixture of the degenerate distribution at 0 (with probability p) and an exponential distribution, that is,

$$F(x) := P\left(\sum_{i=1}^M X_i \leq x\right) = p + (1 - p)(1 - e^{-px}) = 1 - (1 - p)e^{-px}, \quad x \geq 0. \quad (15)$$

When an explicit expression is not possible, the usual approximate evaluation method is by discretizing the summands in (14) and then using recursive methods found in the literature for discrete random sums. By considering (1) as a first discretization method, we have (see [5], page 391)

$$P\left(X_1^{\bullet t} = \frac{k}{t}\right) = \left(\frac{t}{t+1}\right)^k \frac{1}{t+1}, \quad k = 0, 1, \dots \quad (16)$$

Thus, $t \sum_{i=1}^M X_i^{\bullet t}$ is a geometric sum of geometric distributions with parameter $r = (1 + t)^{-1}$. It is easy to check (use LS transforms, for instance) that the distribution of such a random variable is a mixture of the degenerate distribution at 0 (with probability p) and a geometric distribution with parameter $p^* = 1 - (1 - r)(1 - (1 - p)r)^{-1} = 1 - t(t + p)^{-1}$, so that for each $k \in \mathbb{N}$,

$$\begin{aligned} L_t^* F\left(\frac{k}{t}\right) &= P\left(\sum_{i=1}^M X_i^{\bullet t} \leq \frac{k}{t}\right) \\ &= p + (1 - p)(1 - (1 - p^*)^{k+1}) = 1 - (1 - p)\left(\frac{t}{t+p}\right)^{k+1}. \end{aligned} \quad (17)$$

Note that the first equality in (17) follows by recalling (2) and noting that $(\sum_{i=1}^M X_i)^{\bullet t}$ has the same distribution as $\sum_{i=1}^M X_i^{\bullet t}$ (see [16], Proposition 2.1). Actually, a more natural way (in this case) to compute (17) is to evaluate the LS transform of $(\sum_{i=1}^M X_i)^{\bullet t}$ and then apply (1) and (2). However, the previous computations enable easier comparisons with the following method. In fact, one of the most obvious (and widely used) methods to discretize the summands in (14) is by a rounding method. For instance, a rounding down method (we round X_i to $[tX_i]t^{-1}$) yields

$$P\left(\frac{[tX_1]}{t} = \frac{k}{t}\right) = P\left(\frac{k}{t} \leq X_1 < \frac{k+1}{t}\right) = e^{-k/t}(1 - e^{-1/t}), \quad k \in \mathbb{N}. \quad (18)$$

Table 1. Comparison of different approximation methods for (15)

$x = \frac{k}{5}$	$F(\frac{k}{5})$	$L_5^*F(\frac{k}{5})$	$R_5F(\frac{k}{5})$	$W_5F(\frac{k}{5})$
$0 = \frac{0}{5}$	0.1000	0.1176	0.1195	0.1000
$1 = \frac{5}{5}$	0.1856	0.2008	0.2108	0.1848
$5 = \frac{25}{5}$	0.4541	0.4622	0.4907	0.4412
$10 = \frac{50}{5}$	0.6689	0.6722	0.7054	0.6383
$15 = \frac{75}{5}$	0.7992	0.8002	0.8296	0.7576
$20 = \frac{100}{5}$	0.8782	0.8782	0.9014	0.8327
$30 = \frac{150}{5}$	0.9552	0.9548	0.9670	0.9142
$40 = \frac{200}{5}$	0.9835	0.9832	0.9890	0.9524

In this case, $\sum_{i=1}^M [tX_i]$ is a geometric sum of geometric distributions with parameter $r' = 1 - e^{-1/t}$. We denote by R_tF the distribution function of $\sum_{i=1}^M \frac{[tX_i]}{t}$. Using the same arguments as for (17), we obtain for each $k \in \mathbb{N}$ that

$$R_tF\left(\frac{k}{t}\right) = P\left(\sum_{i=1}^M \frac{[tX_i]}{t} \leq \frac{k}{t}\right) = 1 - (1-p) \left(\frac{e^{-1/t}}{1 - (1-p)(1 - e^{-1/t})}\right)^{k+1}. \quad (19)$$

Finally, it would be interesting to compare the previous ‘discretization methods’ with a ‘transform method.’ To this end, we consider the Laplace transform of F in (15) (instead of its LS transform), that is,

$$w_F(\theta) = \int_0^\infty e^{-\theta u} F(u) du = \frac{1}{\theta} - \frac{1-p}{\theta+p}, \quad \theta > 0,$$

and apply the Post–Widder inversion formula (see [10], page 233), defined for $t \in \mathbb{N}^*$ as

$$W_tF(x) = \frac{(-1)^{t-1}}{(t-1)!} \left(\frac{t}{x}\right)^t w_F^{(t-1)}\left(\frac{t}{x}\right) = 1 - \frac{(1-p)t^t}{(px+t)^t}, \quad x \geq 0.$$

In Table 1 (computations with MATLAB) we consider a ‘rough’ discretization interval ($t = 5$), a small p ($p = 0.1$) and present, for different $x = k/5$, the exact values of F (column 2), the L_t^* approximation (column 3), the ‘rounding down’ discretization (column 4) and the Post–Widder inversion (column 5).

As we can see in Table 1, L_5^*F provides a better approximation than R_5F . The intuitive explanation of this fact is that, when approximating $\sum_{i=1}^M X_i$ by $\sum_{i=1}^M X_i^{\bullet t}$, the error in the approximation can be controlled ‘uniformly,’ regardless of the distribution of M (see [16], Theorem 4.3). This effect is obvious when we choose M with a large expected value (our choice of a small p is for this reason – for larger values of p checked, L_5^*F is also better, but the difference is less appreciable). However, if we compare the approximations

Table 2. Comparison of $M_5^{[2]}$ in (10) with $G_5^{[2]}$ in (21)

$x = \frac{k}{5}$	$F(\frac{k}{5})$	$L_5^*F(\frac{k-1}{5})$	$L_{10}^*F(\frac{2k-1}{10})$	$M_5^{[2]}F(\frac{k}{5})$	$G_5^{[2]}F(\frac{k}{5})$
$1 = \frac{5}{5}$	0.1856	0.1848	0.1852	0.1856	0.1856
$5 = \frac{25}{5}$	0.4541	0.4514	0.4528	0.4541	0.4538
$10 = \frac{50}{5}$	0.6689	0.6656	0.6673	0.6689	0.6677
$15 = \frac{75}{5}$	0.7992	0.7962	0.7977	0.7992	0.7975
$20 = \frac{100}{5}$	0.8782	0.8758	0.8770	0.8782	0.8766
$30 = \frac{150}{5}$	0.9552	0.9538	0.9545	0.9552	0.9553
$40 = \frac{200}{5}$	0.9835	0.9829	0.9832	0.9835	0.9854

L_5^*F and W_5F , we see that the last one is better for small values, whereas the first one is better for large values. To explain this fact, it is interesting to point out that W_tF , like L_t^*F , admits the following well-known representation. For a function g defined on $[0, \infty)$, we can write, as in (5) (see [10], pages 220, 223),

$$W_tg(x) = Eg\left(x\frac{S(t)}{t}\right), \quad x > 0. \quad (20)$$

Note that the mean of the ‘random points’ defining W_t in (20) is $E(xt^{-1}S(t)) = x$, whereas for L_t^* in (5), we have $E(t^{-1}S([tx] + 1)) = t^{-1}([tx] + 1)$. This means that W_t is centered at x , whereas L_t^* is ‘biased’. The benefits of this property for W_t are observed at small values in Table 1. However, we have $\text{Var}(xt^{-1}S(t)) = t^{-1}x^2$, whereas $\text{Var}(t^{-1}S([tx] + 1)) = t^{-2}([tx] + 1)$, the latter being of order $t^{-1}x$, as $t \rightarrow \infty$. The greater variability of the random variables defining W_t for greater values of x produces an undesired effect in the approximation.

We now show the improvement in the approximation which occurs when using $M_t^{[2]}$, as defined in (10), instead of L_t^* . In Table 2 below ($t = 5$), we compare $M_5^{[2]}F$ (column 5) with Richardson extrapolation for W_tF (or Stehfest enhancement of order two for the Post–Widder formula – see [1], page 40), that is,

$$G_t^{[2]}F(x) := 2W_{2t}F(x) - W_tF(x), \quad x > 0. \quad (21)$$

As we can see, $M_5^{[2]}F$ provides us with an exact value up to a four decimal places, whereas $G_5^{[2]}F$ does not achieve this accuracy.

3. Error bounds for the approximation

Let $g \in \mathcal{D}$, as defined in (7). Our first aim is to give bounds of $\|L_t^{[2]}g - g\|$ in terms of $\|x^2 g^{iv}(x)\|$. To this end, we will use the following as ‘test function’:

$$\phi(x) = \begin{cases} 0, & \text{if } x = 0, \\ \frac{x^2}{2} \left(\frac{3}{2} - \log(x) \right), & \text{otherwise.} \end{cases} \quad (22)$$

Observe that $\phi \in \mathcal{D}$. In fact, by elementary calculus,

$$\phi'(x) = x(1 - \log x), \quad \phi''(x) = -\log x, \quad \phi'''(x) = -\frac{1}{x} \quad \text{and} \quad \phi^{iv}(x) = \frac{1}{x^2}. \quad (23)$$

In the next lemma, we make an explicit computation of $L_t\phi(x)$ in terms of the Ψ (or digamma) function. This function is defined as (see [3], page 258)

$$\Psi(x) := \frac{d}{dx} \log(\Gamma(x)) = \frac{1}{\Gamma(x)} \int_0^\infty \log u e^{-u} u^{x-1} du, \quad x > 0, \quad (24)$$

and, therefore, using the last equality, we have the following probabilistic expression of the psi function in terms of the gamma process:

$$\Psi(x) = E \log S(x), \quad x > 0. \quad (25)$$

We will use the following property of this function (see [3], page 258):

$$\Psi(x+1) = \frac{1}{x} + \Psi(x). \quad (26)$$

Lemma 3.1. *Let ϕ be defined as in (22) and let $L_t, t > 0$, be defined as in (4). We have that*

$$L_t\phi(x) = \frac{1}{2t^2} \left(\frac{3(tx)^2}{2} - \frac{tx}{2} - 1 + tx(tx+1)(-\Psi(tx) + \log(t)) \right), \quad x > 0. \quad (27)$$

Proof. Let $t > 0$ and $x > 0$ be fixed. First, using elementary calculus, (4) and (26), we can write

$$\begin{aligned} L_t\phi(x) &= E \frac{S(tx)^2}{2t^2} \left(\frac{3}{2} - \log \left(\frac{S(tx)}{t} \right) \right) \\ &= \frac{1}{2t^2} \frac{1}{\Gamma(tx)} \int_0^\infty u^2 \left(\frac{3}{2} - \log \left(\frac{u}{t} \right) \right) e^{-u} u^{tx-1} du \\ &= \frac{(tx)(tx+1)}{2t^2} \frac{1}{\Gamma(tx+2)} \int_0^\infty \left(\frac{3}{2} - \log \left(\frac{u}{t} \right) \right) e^{-u} u^{tx+1} du \\ &= \frac{(tx)(tx+1)}{2t^2} \left(\frac{3}{2} - E \log \left(\frac{S(tx+2)}{t} \right) \right). \end{aligned} \quad (28)$$

Therefore, using (25), we can write

$$L_t\phi(x) = \frac{(tx)(tx+1)}{2t^2} \left(\frac{3}{2} - \Psi(tx+2) + \log(t) \right). \quad (29)$$

Now, using (26) twice, we have

$$\Psi(tx+2) = \frac{2(tx)+1}{tx(tx+1)} + \Psi(tx). \quad (30)$$

By (29), (30), we obtain

$$L_t\phi(x) = \frac{(tx)(tx+1)}{2t^2} \left(\frac{3}{2} - \frac{2(tx)+1}{tx(tx+1)} - \Psi(tx) + \log(t) \right).$$

The result follows using elementary algebra in the expression above. \square

In the next lemma, we will study the approximation properties of $L_t\phi$ to ϕ . We will make use of the following inequalities for the psi function:

$$\frac{1}{2x} \leq \log(x) - \Psi(x) \leq \frac{1}{x}, \quad x > 0; \quad (31)$$

$$\log(x) - \Psi(x) - \frac{1}{2x} \leq \frac{1}{12x^2}, \quad x > 0. \quad (32)$$

We can find (31) in [7], page 374, whereas (32) is an immediate consequence of the fact that the function

$$\Psi(x) - \log(x) + \frac{1}{2x} + \frac{1}{12x^2}$$

is completely monotonic (see [15], page 304) and thus non-negative.

Lemma 3.2. *Let ϕ be as defined in (22) and let $L_t, t > 0$, be as defined in (4). We have*

$$\left\| L_t\phi(x) - \phi(x) + \frac{x \log x}{2t} + \frac{1}{3t^2} \right\| \leq \frac{3}{8t^2}. \quad (33)$$

Proof. Let $x > 0$ and $t > 0$ be fixed. First of all, we can write

$$\phi(x) = \frac{1}{2t^2} \left(\frac{3(tx)^2}{2} - (tx)^2 \log(tx) + (tx)^2 \log(t) \right). \quad (34)$$

On the other hand,

$$\frac{x \log x}{2t} + \frac{1}{3t^2} = \frac{1}{2t^2} \left((tx) \log tx - (tx) \log t + \frac{2}{3} \right). \quad (35)$$

Therefore, using Lemma 3.1, (34) and (35), we can write

$$\begin{aligned} L_t \phi(x) - \phi(x) + \frac{x \log x}{2t} + \frac{1}{3t^2} \\ = \frac{1}{2t^2} \left(-\frac{tx}{2} - 1 - (tx)^2 \Psi(tx) - (tx) \Psi(tx) + (tx)^2 \log(tx) + (tx) \log(tx) + \frac{2}{3} \right) \\ = \frac{1}{2t^2} \left((tx)^2 \left(\log(tx) - \Psi(tx) - \frac{1}{2(tx)} \right) + tx(\log(tx) - \Psi(tx)) - \frac{1}{3} \right). \end{aligned} \quad (36)$$

By (31), we have that $1/2 \leq x(\log(x) - \Psi(x)) \leq 1, x > 0$, and thus

$$\frac{1}{6} \leq tx(\log(tx) - \Psi(tx)) - \frac{1}{3} \leq \frac{2}{3}. \quad (37)$$

Thus, using (36), (37) and (32), we obtain (33). \square

We are now in a position to state the following.

Theorem 3.1. *Let $g \in \mathcal{D}$, as defined in (7) and let $L_t^{[2]}, t > 0$, be as defined in (6). We have*

$$|L_t^{[2]} g(x) - g(x)| \leq \frac{1}{6t^2} \|xg'''(x)\| + \frac{9}{16t^2} \|x^2 g^{iv}(x)\|.$$

Proof. We will first see that $g \in \mathcal{D}$ implies that

$$\|xg'''(x)\| \leq \|x^2 g^{iv}(x)\| < \infty. \quad (38)$$

To begin with, the fact that $\|x^2 g^{iv}(x)\| < \infty$ implies that $\lim_{x \rightarrow \infty} x^{1+\alpha} g^{iv}(x) = 0$ for all $0 < \alpha < 1$. By L'Hôpital's rule, we also have that $\lim_{x \rightarrow \infty} x^\alpha g'''(x) = 0$, thus concluding that $\lim_{x \rightarrow \infty} g'''(x) = 0$. Using this fact, we can write

$$g'''(x) = - \int_x^\infty g^{iv}(u) \, du,$$

which implies easily (38) as

$$|xg'''(x)| \leq x \int_x^\infty \frac{|u^2 g^{iv}(u)|}{u^2} \, du \leq \|x^2 g^{iv}(x)\|.$$

Now, let $t > 0$ and let L_t be as in (4). As a previous step, we will prove that

$$\left| L_t g(x) - g(x) - \frac{xg''(x)}{2t} - \frac{xg'''(x)}{3t^2} \right| \leq \frac{3}{8t^2} \|x^2 g^{iv}(x)\|, \quad x > 0. \quad (39)$$

To this end, let $x > 0$. Using a Taylor series expansion of the random point $u = S(tx)/t$ around x and taking into account that $E(S(x) - x) = 0$, $E(S(x) - x)^2 = x$ and $E(S(x) -$

$x)^3 = 2x$, we can write

$$\begin{aligned}
L_t g(x) - g(x) &= E g\left(\frac{S(tx)}{t}\right) - g(x) \\
&= \frac{E(S(tx) - tx)^2}{2t^2} g''(x) + \frac{E(S(tx) - tx)^3}{6t^3} g'''(x) \\
&\quad + \frac{1}{6} E \int_x^{S(tx)/t} g^{iv}(\theta) \left(\frac{S(tx)}{t} - \theta\right)^3 d\theta \\
&= \frac{xg''(x)}{2t} + \frac{xg'''(x)}{3t^2} + \frac{1}{6} E \int_x^{S(tx)/t} g^{iv}(\theta) \left(\frac{S(tx)}{t} - \theta\right)^3 d\theta.
\end{aligned} \tag{40}$$

Then, using (40), we get the bound

$$\begin{aligned}
&\left| L_t g(x) - g(x) - \frac{xg''(x)}{2t} - \frac{xg'''(x)}{3t^2} \right| \\
&= \frac{1}{6} \left| E \int_x^{S(tx)/t} g^{iv}(\theta) \left(\frac{S(tx)}{t} - \theta\right)^3 d\theta \right| \\
&\leq \frac{\|x^2 g^{iv}(x)\|}{6} E \int_{\min(x, S(tx)/t)}^{\max(x, S(tx)/t)} \left| \frac{S(tx)}{t} - \theta \right|^3 \frac{1}{\theta^2} d\theta \\
&= \frac{\|x^2 g^{iv}(x)\|}{6} E \int_x^{S(tx)/t} \left(\frac{S(tx)}{t} - \theta\right)^3 \frac{1}{\theta^2} d\theta.
\end{aligned} \tag{41}$$

Let $\phi(\cdot)$ be as in (22). Using (40) and (23), we have

$$L_t \phi(x) - \phi(x) + \frac{x \log x}{2t} + \frac{1}{3t^2} = \frac{1}{6} E \int_x^{S(tx)/t} \left(\frac{S(tx)}{t} - \theta\right)^3 \frac{1}{\theta^2} d\theta. \tag{42}$$

Then, by (41) and (42), we can write

$$\left| L_t g(x) - g(x) - \frac{xg''(x)}{2t} - \frac{xg'''(x)}{3t^2} \right| \leq \|x^2 g^{iv}(x)\| \left\| L_t \phi(x) - \phi(x) + \frac{x \log x}{2t} + \frac{1}{3t^2} \right\|.$$

Thus, (39) follows by applying Lemma 3.2.

Observe that in (39), the only term of order $1/t$ is the one involving the second derivative. By means of the operator $L_t^{[2]}$, as defined in (6), this term is eliminated. In fact, using (39), we have

$$\begin{aligned}
L_t^{[2]} g(x) - g(x) &= 2(L_{2t} g(x) - g(x)) - (L_t g(x) - g(x)) \\
&= 2 \left(L_{2t} g(x) - g(x) - \frac{x}{4t} g''(x) - \frac{x}{12t^2} g'''(x) \right)
\end{aligned}$$

$$\begin{aligned}
& - \left(L_t g(x) - g(x) - \frac{x}{2t} g''(x) - \frac{x}{3t^2} g'''(x) \right) - \frac{x}{6t^2} g'''(x) \\
& \leq \frac{1}{6t^2} \|xg'''(x)\| + \frac{9}{16t^2} \|x^2 g^{iv}(x)\|.
\end{aligned} \tag{43}$$

This completes the proof of Theorem 3.1. \square

Finally, in the next result, we study the approximation properties of $M_t^{[2]}$.

Theorem 3.2. *Let $g \in \mathcal{D}_1$, as defined in (9) and let $M_t^{[2]}, t > 0$, be as defined in (8). We have*

$$\|M_t^{[2]}g - g\| \leq \frac{1}{8t^2} \|g''(x)\| + \frac{1}{6t^2} \|xg'''(x)\| + \frac{9}{16t^2} \|x^2 g^{iv}(x)\|.$$

Proof. First, note that $g \in \mathcal{D}_1$ implies that $\|xg'''(x)\| < \infty$, thanks to (38). Now, let $t > 0$ and $x > 0$ be fixed. We write

$$\begin{aligned}
M_t^{[2]}g(x) - g(x) &= (tx - [tx]) \left(L_t^{[2]}g\left(\frac{[tx] + 1}{t}\right) - g\left(\frac{[tx] + 1}{t}\right) \right) \\
&\quad + ([tx] + 1 - tx) \left(L_t^{[2]}g\left(\frac{[tx]}{t}\right) - g\left(\frac{[tx]}{t}\right) \right) \\
&\quad + (tx - [tx]) \left(g\left(\frac{[tx] + 1}{t}\right) - g(x) \right) \\
&\quad + ([tx] + 1 - tx) \left(g\left(\frac{[tx]}{t}\right) - g(x) \right).
\end{aligned} \tag{44}$$

Using the usual expansion

$$|g(y) - g(x) - (y - x)g'(x)| \leq \frac{(y - x)^2}{2} \|g''\| \tag{45}$$

and taking into account that

$$\begin{aligned}
& (tx - [tx]) \left(g\left(\frac{[tx] + 1}{t}\right) - g(x) \right) + ([tx] + 1 - tx) \left(g\left(\frac{[tx]}{t}\right) - g(x) \right) \\
&= (tx - [tx]) \left(g\left(\frac{[tx] + 1}{t}\right) - g(x) - \frac{[tx] + 1 - tx}{t} g'(x) \right) \\
&\quad + ([tx] + 1 - tx) \left(g\left(\frac{[tx]}{t}\right) - g(x) - \frac{[tx] - tx}{t} g'(x) \right),
\end{aligned} \tag{46}$$

we obtain from the above expression and (45) that

$$\begin{aligned}
& \left| (tx - [tx]) \left(g\left(\frac{[tx] + 1}{t}\right) - g(x) \right) + ([tx] + 1 - tx) \left(g\left(\frac{[tx]}{t}\right) - g(x) \right) \right| \\
& \leq \left((tx - [tx]) \frac{([tx] + 1 - tx)^2}{2t^2} + ([tx] + 1 - tx) \frac{([tx] - tx)^2}{2t^2} \right) \|g''\| \\
& = \frac{(tx - [tx])([tx] + 1 - tx)}{2t^2} \|g''\| \leq \frac{1}{8t^2} \|g''\|,
\end{aligned} \tag{47}$$

the last inequality holding since for each $k \in \mathbb{N}$, the supremum of $(u - k)(k + 1 - u)$, $k \leq u \leq k + 1$, is attained at $u = k + 1/2$. On the other hand, taking into account Theorem 3.1, we have

$$\begin{aligned}
& \left| (tx - [tx]) \left(L_t^{[2]} g\left(\frac{[tx] + 1}{t}\right) - g\left(\frac{[tx] + 1}{t}\right) \right) \right. \\
& \quad \left. + ([tx] + 1 - tx) \left(L_t^{[2]} g\left(\frac{[tx]}{t}\right) - g\left(\frac{[tx]}{t}\right) \right) \right| \\
& \leq \|L_t^{[2]} g - g\| \leq \frac{1}{6t^2} \|xg'''(x)\| + \frac{9}{16t^2} \|x^2 g^{iv}(x)\|.
\end{aligned} \tag{48}$$

The result follows by (44), (47) and (48). \square

4. Application to gamma distributions

In this section, we will study the case of gamma distributions, that is, distributions with density functions as given in (3). It is not hard to see that these distributions are in the class \mathcal{D}_1 , for a shape parameter $p = 1$ or $p \geq 2$, and, therefore, we are a position of apply Theorem 3.2. The aim of this section is to show that, in fact, the bounds in this theorem can be uniformly bounded on the shape parameter, which will be an advantage when dealing with mixtures of these distributions. From now on, we define

$$f_p(x) := \begin{cases} \frac{e^{-x} x^{p-1}}{\Gamma(p)}, & x > 0, \text{ if } p \in \mathbb{R} \setminus \{0, -1, -2, \dots\}, \\ 0, & x > 0, \text{ if } p \in \{0, -1, -2, \dots\}. \end{cases} \tag{49}$$

The ‘odd’ definition of f_p for $p \in \{0, -1, -2, \dots\}$ is for notational convenience in (51). For $p > 0$, the function above is the density of a gamma random variable as in (3), with scale parameter $a = 1$. Results for another scale parameter will follow by a change of scale (see Proposition 5.1 below). First, we will consider the case $p = 1$, that is, an exponential random variable. As the distribution function of this random variable presents no computational problems, it makes no sense to approximate it. However, when we consider the problem of approximating a general mixture of Gamma distributions, the exponential distribution could be a component.

Lemma 4.1. *Let $F(x) = 1 - e^{-x}$, $x \geq 0$. For $t > 0$, let $M_t^{[2]}F$ be as defined in (8). We have that*

$$\|M_t^{[2]}F - F\| \leq \left(\frac{1}{8} + \frac{1}{6e} + \frac{9}{4e^2}\right) \frac{1}{t^2}.$$

Proof. First of all, note that $|F^{(k)}(x)| = e^{-x}$ and that $\sup_{x \geq 0} x^k e^{-x} = k^k e^{-k}$, $k = 1, 2, \dots$. Thus, we have

$$\|F''\| = 1, \quad \|xF'''(x)\| = e^{-1} \quad \text{and} \quad \|x^2 F^{iv}(x)\| = 2^2 e^{-2}. \quad (50)$$

The conclusion follows by taking into account Theorem 3.2. \square

We will now deal with the case $p \geq 2$ in (49). The two following lemmas will be useful in order to bound the derivatives of this density. For the sake of brevity, they are stated without proof (only elementary calculus is required). For the proofs, we refer the interested reader to [17], a preliminary version of this paper (available online).

Lemma 4.2. *Let $f_p(\cdot)$, $p > 0$, be as defined in (49). We have, for all $n \in \mathbb{N}$,*

$$\begin{aligned} \frac{d^n}{dx^n} f_p(x) &= \frac{e^{-x} x^{p-n-1}}{\Gamma(p)} \sum_{i=0}^n \binom{n}{i} (-1)^i \left(\prod_{j=1}^{n-i} (p-j) \right) x^i \\ &= \sum_{i=0}^n \binom{n}{i} (-1)^i f_{p-n+i}(x), \quad x > 0, \end{aligned} \quad (51)$$

in which $\prod_{j=1}^0 (p-j) = 1$.

Next, we formulate a technical lemma in which we define certain decreasing functions which will be used to bound the weighted derivatives of f_p .

Lemma 4.3. *We have:*

(i) *the function*

$$g_1(p) := \frac{1}{\Gamma(p)} e^{-(p-1)} (p-1)^{p-1}, \quad p > 1 \quad (52)$$

($g_1(1) = 1$), *is decreasing in p ;*

(ii) *the function*

$$g_2(p) := \frac{1}{\Gamma(p)} e^{-(p-1/2+1/2\sqrt{4p-3})} \left(p - \frac{1}{2} + \frac{1}{2}\sqrt{4p-3} \right)^{p-1/2}, \quad p \geq 1, \quad (53)$$

is decreasing in p ;

(iii) the function

$$g_3(p) := \frac{1}{\Gamma(p)} e^{-(p-1-\sqrt{p-1})} (\sqrt{p-1}-1)^{p-2} (\sqrt{p-1})^{p-1}, \quad p > 2 \quad (54)$$

($g_3(2) = 1$), is decreasing in p ;

(iv) the function

$$g_4(p) := \frac{1}{\Gamma(p)} e^{-(p-\sqrt{3p-2})} (p-\sqrt{3p-2})^{p-2} (\sqrt{3p-2}-1)^3, \quad p > 2 \quad (55)$$

($g_4(2) = 1$), is decreasing in p .

In the following result, we get bounds of the quantities required in Theorem 3.2, depending on the shape parameter p , but also decreasing on p .

Lemma 4.4. Let f_p be as in (49) and $g_i, i = 1, 2, 3, 4$, be as in Lemma 4.3. We have:

- (i) $\sup_{x \geq 0} |f_p(x)| = g_1(p), \quad p \geq 1;$
- (ii) $\sup_{x \geq 0} |x f'_p(x)| = g_2(p), \quad p \geq 1;$
- (iii) $\sup_{x \geq 0} |f'_p(x)| = g_3(p), \quad p \geq 2;$
- (iv) $\sup_{x \geq 0} |x f''_p(x)| \leq \max\{g_1(p-1), g_2(p-1)\}, \quad p \geq 2;$
- (v) $\sup_{x \geq 0} |x^2 f'''_p(x)| \leq g_4(p) + 3g_2(p-1) + g_1(p-1), \quad p \geq 2.$

Proof. To show part (i), it is clear that, for $p \geq 1$,

$$\sup_{x \geq 0} f_p(x) = f_p(p-1) = \frac{e^{-(p-1)}(p-1)^{p-1}}{\Gamma(p)}$$

and (i) follows by recalling (52). To show part (ii), we have (see [16], Remark 3.2 and Lemma 5.2)

$$\sup_{x \geq 0} |x f'_p(x)| = \frac{1}{\Gamma(p)} \left(p - \frac{1}{2} + \frac{1}{2} \sqrt{4p-3} \right)^{p-1/2} e^{-p-1/2+1/2\sqrt{4p-3}}, \quad p > 1, \quad (56)$$

and (ii) follows by recalling (53). To show part (iii), by (51), we have for $p \geq 2$ that

$$f'_p(x) = \frac{1}{\Gamma(p)} e^{-x} x^{p-2} (p-1-x), \quad x > 0, \quad (57)$$

$$f''_p(x) = \frac{1}{\Gamma(p)} e^{-x} x^{p-3} ((p-1)(p-2) - 2(p-1)x + x^2), \quad x > 0, \quad (58)$$

and it can be easily checked that the zeros of $f_p''(x)$ are $p_1 := p - 1 - \sqrt{p-1}$ and $p_2 := p - 1 + \sqrt{p-1}$. Therefore, $|f_p'(x)|$ must attain its maximum value at either p_1 or p_2 . Actually, p_1 corresponds to the maximum. To show that, we will see that

$$\frac{f_p'(p_1)}{|f_p'(p_2)|} = e^{2\sqrt{p-1}} \left(\frac{\sqrt{p-1}-1}{\sqrt{p-1}+1} \right)^{p-2} \geq 1, \quad p \geq 2. \quad (59)$$

To show the last inequality in (59), taking logarithms, we will prove that

$$r_1(p) := 2\sqrt{p-1} + (p-2)(\log(\sqrt{p-1}-1) - \log(\sqrt{p-1}+1)) \geq 0, \quad p > 2. \quad (60)$$

Define

$$\rho_1(b) := \frac{2b}{b^2-1} + (\log(b-1) - \log(b+1)), \quad b > 1.$$

Note that

$$r_1(p) = (p-2)\rho_1(\sqrt{p-1}), \quad p > 2. \quad (61)$$

We will first prove that

$$\rho_1(b) \geq 0, \quad b > 1. \quad (62)$$

To show (62), it is readily seen that $\rho_1'(b) = -4(b^2-1)^{-2}$, $b > 1$, so that ρ_1 is decreasing. As $\lim_{b \rightarrow \infty} \rho_1(b) = 0$, we have (62). This implies also (60), recalling (61). Therefore, we conclude that

$$\sup_{x>0} |f_p'(x)| = f_p'(p_1) = \frac{1}{\Gamma(p)} e^{-(p-1-\sqrt{p-1})} (\sqrt{p-1}-1)^{p-2} (\sqrt{p-1})^{p-1}, \quad (63)$$

which, together with (54), shows (iii).

To show part (iv), note that by using (51), we can write $f_p'(x) = f_{p-1}(x) - f_p(x)$ and, therefore,

$$x f_p''(x) = x f_{p-1}'(x) - x f_p'(x), \quad x > 0, p \geq 2. \quad (64)$$

On the other hand, we see in (58) that $f_{p-1}'(x)$ and $f_p'(x)$ have the same sign for $0 < x < p-2$ and $p-1 < x < \infty$ and, therefore, using part (ii) and Lemma 4.3(i), we can write

$$\sup_{x \notin [p-2, p-1]} |x f_p''(x)| \leq \max(g_2(p-1), g_2(p)) = g_2(p-1). \quad (65)$$

On the other hand, we have, by (58),

$$x f_p''(x) = \frac{1}{\Gamma(p)} e^{-x} x^{p-2} ((p-1)(p-2) - 2(p-1)x + x^2). \quad (66)$$

Using the above expression and taking into account that, for $p-2 \leq x \leq p-1$,

$$e^{-x} x^{(p-2)} \leq e^{-p-2} (p-2)^{p-2} \quad \text{and} \quad |(p-1)(p-2) - 2(p-1)x + x^2| = p-1, \quad (67)$$

the last inequality holds as $|(p-1)(p-2) - 2(p-1)x + x^2|, p-2 \leq x \leq p-1$, attains its maximum value at $p-1$. From (66) and (67), we conclude that

$$\sup_{x \in [p-2, p-1]} |xf_p''(x)| \leq \frac{1}{\Gamma(p)} e^{-(p-2)} (p-2)^{p-2} (p-1) = g_1(p-1), \quad (68)$$

where the last inequality follows by recalling (52). Thus, (65) and (68) conclude the proof of part (iv). To show part (v), let $p \geq 2$. First, we have, by (51),

$$\begin{aligned} f_p'''(x) &= f_{p-3}(x) - 3f_{p-2}(x) + 3f_{p-1}(x) - f_p(x) \\ &= \frac{e^{-x}x^{p-4}}{\Gamma(p)} ((p-1)(p-2)(p-3) - 3(p-1)(p-2)x + 3(p-1)x^2 - x^3) \\ &= \frac{e^{-x}x^{p-4}}{\Gamma(p)} ((p-1-x)^3 + 3(p-1)(x-(p-2)) - (p-1)), \quad x > 0. \end{aligned} \quad (69)$$

Therefore, if we call

$$h_p(x) := \frac{e^{-x}x^{p-2}}{\Gamma(p)} (p-1-x)^3, \quad x > 0,$$

we have, recalling (57),

$$\begin{aligned} x^2 f_p'''(x) &= \frac{e^{-x}x^{p-2}}{\Gamma(p)} ((p-1-x)^3 - 3(p-1)(x-(p-2)) - (p-1)) \\ &= h_p(x) + 3xf_{p-1}'(x) - f_{p-1}(x), \quad x \geq 0. \end{aligned} \quad (70)$$

We will firstly see that

$$\sup_{x \geq 0} |h_p(x)| = g_4(p) \quad (71)$$

with $g_4(\cdot)$ as defined in (55). Note that

$$h_p'(x) = \frac{e^{-x}x^{p-3}}{\Gamma(p)} (p-1-x)^2 (x^2 - 2px + (p-1)(p-2)), \quad x > 0.$$

The maximum value of $|h_p|$ will be attained at the roots of the last polynomials, being $p_1 := p + \sqrt{3p-2}$ and $p_2 := p - \sqrt{3p-2}$. To check which value attains the maximum, define $u := \sqrt{3p-2}$. Note that $p_1 = (u+1)(u+2)/3$ and $p_2 = (u-1)(u-2)/3$. Then, with this notation, we will prove that

$$\frac{|h_p(p_2)|}{|h_p(p_1)|} = e^{2u} \left(\frac{(u-1)(u-2)}{(u+1)(u+2)} \right)^{(u^2-4)/3} \left(\frac{u-1}{u+1} \right)^3 \geq 1, \quad u > 2. \quad (72)$$

To show the last inequality in (72), taking logarithms, we will show that

$$\rho_2(u) := 2u + \frac{u^2 - 4}{3} \log\left(\frac{(u-1)(u-2)}{(u+1)(u+2)}\right) + 3 \log\left(\frac{u-1}{u+1}\right) \geq 0, \quad u > 2. \quad (73)$$

Note that

$$\begin{aligned} \rho'_2(u) &= 2 + \frac{2u}{3} \log\left(\frac{(u-1)(u-2)}{(u+1)(u+2)}\right) + \frac{u^2 - 4}{3} \left(\frac{1}{u-1} + \frac{1}{u-2} - \frac{1}{u+1} - \frac{1}{u+2}\right) \\ &\quad + 3 \left(\frac{1}{u-1} - \frac{1}{u+1}\right) \\ &= \frac{4u^2}{u^2 - 1} + \frac{2u}{3} \log\left(\frac{(u-1)(u-2)}{(u+1)(u+2)}\right), \quad u > 2. \end{aligned}$$

We will show that $\rho'_2(u) \leq 0$, $u > 2$. In fact,

$$\frac{d}{du} \frac{3}{2u} \rho'_2(u) = \frac{36}{(u+1)^2(u-1)^2(u^2-4)^2} \geq 0, \quad u > 2,$$

and then $3(2u)^{-1} \rho'_2(u)$ is increasing. As $\lim_{u \rightarrow \infty} 3(2u)^{-1} \rho'_2(u) = 0$, we conclude that $3(2u)^{-1} \times \rho'_2(u) \leq 0$ and thus that $\rho'_2(u) \leq 0$. Therefore, $\rho_2(u)$ is decreasing. This, together with the fact that $\lim_{u \rightarrow \infty} \rho_2(u) = 0$, proves (73) and therefore (72). Then, $\|h_p\| = h_p(p_2) = g_4(p)$, thus proving (71). The proof of part (iv) now follows easily by recalling (70) and using (71) and parts (i) and (ii). \square

As an immediate consequence of Theorem 3.2 and Lemma 4.4, we have the following corollary.

Corollary 4.1. *Let F_p be a gamma distribution with shape parameter $p \geq 2$, that is, whose density function is given by (49). Let $M_t^{[2]}$, $t > 0$, be defined as in (8). We have*

$$\|M_t^{[2]} F_p - F_p\| \leq \left(\frac{17}{12} + \frac{27}{16e}\right) \frac{1}{t^2} \approx \frac{2.0375}{t^2}.$$

Proof. Let $p \geq 2$ be fixed. The result is an immediate consequence of Theorem 3.2, as $F'_p = f_p$, as defined in (49). Therefore, by Lemma 4.4(iii) and Lemma 4.3(ii), we have that

$$\|F_p''\| = \|f_p'\| = g_3(p) \leq g_3(2) = 1. \quad (74)$$

On the other hand, we see that by Lemma 4.3(i), we have that

$$g_1(p-1) \leq g_1(1) = 1 \quad \text{and} \quad g_2(p-1) \leq g_2(1) = e^{-1}, \quad p \geq 2. \quad (75)$$

Thus, using the above inequalities and Lemma 4.4(iv), we have

$$\|x F_p'''(x)\| = \|x f_p''(x)\| \leq 1. \quad (76)$$

Finally by Lemma 4.4(v), Lemma 4.3(iv) and (75), we have

$$\|x^2 F_p^{iv}(x)\| = \|x^2 f_p'''(x)\| \leq g_4(2) + 3g_2(1) + g_1(1) = 2 + 3e^{-1}. \quad (77)$$

Using (74), (76), (77) and Theorem 3.2, we obtain the result. This completes the proof of Corollary 4.1. \square

5. Applications to mixtures of Erlang distributions and phase-type distributions

In this section we apply the results from the previous section to mixtures of Erlang distributions and to random sums of thereof. In order to undertake this study for an arbitrary scale parameter, we need the following result which shows the behavior of $M_t^{[2]}F$ under changes of scale.

Proposition 5.1. *Let X be a random variable with distribution function F . For a given $c > 0$, denote by F^c the distribution function of cX . Let $M_t^{[2]}F$ and $M_t^{[2]}F^c$, $t > 0$, be the respective approximations for F and F^c , as defined in (8). We have that*

$$M_t^{[2]}F^c(x) = M_{ct}^{[2]}F(x/c), \quad x \geq 0. \quad (78)$$

Therefore,

$$\|M_t^{[2]}F^c - F^c\| = \|M_{ct}^{[2]}F - F\|. \quad (79)$$

Proof. Let $t > 0$ and $c > 0$ be fixed. First, we will see that

$$M_t^{[2]}F^c\left(\frac{k}{t}\right) = M_{ct}^{[2]}F\left(\frac{k}{ct}\right), \quad k \in \mathbb{N}, \quad (80)$$

and, therefore, (78) is satisfied for points in the set $k/t, k \in \mathbb{N}$. To this end, we use (12) and (6), and take into account that

$$F^c(x) = F(x/c), \quad x \geq 0, \quad (81)$$

to write, for all $k \in \mathbb{N}$,

$$\begin{aligned} M_t^{[2]}F^c\left(\frac{k}{t}\right) &= 2EF^c\left(\frac{S(2k)}{2t}\right) - EF^c\left(\frac{S(k)}{t}\right) \\ &= 2EF\left(\frac{S(2k)}{2ct}\right) - EF\left(\frac{S(k)}{ct}\right) = M_{ct}^{[2]}F\left(\frac{k}{ct}\right), \end{aligned} \quad (82)$$

thus proving (80). For a general $x > 0$, we use (8) and (80), to see that

$$\begin{aligned} M_t^{[2]} F^c(x) &= (tx - [tx]) M_t^{[2]} F^c\left(\frac{[tx] + 1}{t}\right) + ([tx] + 1 - tx) M_t^{[2]} F^c\left(\frac{[tx]}{t}\right) \\ &= (tx - [tx]) M_{ct}^{[2]} F\left(\frac{[tx] + 1}{ct}\right) + ([tx] + 1 - tx) M_{ct}^{[2]} F\left(\frac{[tx]}{ct}\right) = M_{ct}^{[2]} F\left(\frac{x}{c}\right), \end{aligned}$$

the last inequality being trivial as $tx = (ct)(x/c)$. This concludes the proof of (78). Finally, (79) follows easily from (78) and (81), as we have

$$\sup_{x>0} |M_t^{[2]} F^c(x) - F^c(x)| = \sup_{x>0} |M_{ct}^{[2]} F(x/c) - F(x/c)|.$$

This concludes the proof of Proposition 5.1. \square

As an application of the results in the previous section, we will consider the class of (possibly infinite) mixtures of Erlang distributions recently studied by Willmot and Woo (see [19]). More specifically, let $F_{(a,j)}$, $a > 0, j \in \mathbb{N}^*$, be the distribution function corresponding to the density $f_{(a,j)}$ given in (3) (an Erlang j distribution with scale parameter a). We will consider a finite number of scale parameters arranged in increasing order ($0 < a_1 < \dots < a_n$) and a set of non-negative numbers $p_{ij}, i = 1, \dots, n, j = 0, 1, 2, \dots$, such that $\sum_{i=1}^n \sum_{j=1}^{\infty} p_{ij} = p \leq 1$, and define the class of distribution functions $\mathcal{ME}(a_1, \dots, a_n)$ given as

$$F(x) = (1 - p) + \sum_{i=1}^n \sum_{j=1}^{\infty} p_{ij} F_{a_i,j}(x), \quad x \geq 0 \quad (83)$$

(we consider a slight modification of the class in [19], page 103, as we allow the point mass at 0 with probability $1 - p$). Based on [19], page 103, we can alternatively write (83) by using only the maximum of the scale parameters, that is,

$$F(x) = (1 - p) + \sum_{j=1}^{\infty} p_j F_{a_n,j}(x), \quad x \geq 0. \quad (84)$$

Moreover, the class (84) is a wide class containing many of the distributions considered in applied probability, such as (obviously) finite mixtures of Erlang distributions, but also the class of phase-type distributions (see Proposition 5.3 below). Every random variable having a representation as in (83) can be approximated by means of $M_t^{[2]}$, as shown in the following result.

Proposition 5.2. *Let F be a distribution function of the form $\mathcal{ME}(a_1, \dots, a_n)$, $0 < a_1 < \dots < a_n$, as in (83). Let $M_t^{[2]}, t > 0$, be defined as in (8). We have*

$$\|M_t^{[2]} F - F\| \leq \left(\frac{17}{12} + \frac{27}{16e} \right) \frac{\sum_{i=1}^n (\sum_{j=1}^{\infty} p_{ij}) a_i^2}{t^2}. \quad (85)$$

Proof. Let $t > 0$ and $0 < a_1 < \dots < a_n$ be fixed. The linearity of $M_t^{[2]}$ yields

$$M_t^{[2]}F(x) = (1-p) + \sum_{i=1}^n \sum_{j=1}^{\infty} p_{ij} M_t^{[2]}F_{a_i,j}(x), \quad x \geq 0. \quad (86)$$

By Corollary 4.1, we can write, for a scale parameter 1,

$$\|M_t^{[2]}F_{1,j} - F_{1,j}\| \leq \left(\frac{17}{12} + \frac{27}{16e}\right) \frac{1}{t^2}, \quad j = 2, 3, \dots \quad (87)$$

Moreover, using Lemma 4.1, we have

$$\|M_t^{[2]}F_{1,1} - F_{1,1}\| \leq \left(\frac{1}{2} + \frac{1}{6e} + \frac{9}{4e^2}\right) \frac{1}{t^2} \leq \left(\frac{17}{12} + \frac{27}{16e}\right) \frac{1}{t^2}. \quad (88)$$

Now, let the general scale parameters be $a_i, i = 1, \dots, n$. We use the fact that given X , a gamma random variable of scale parameter 1, X/a_i is a gamma random variable of scale parameter a_i , and, therefore, using Proposition 5.1, (87) and (88), we have for each $a_i, i = 1, \dots, n$, and $j \in \mathbb{N}^*$,

$$\|M_t^{[2]}F_{a_i,j} - F_{a_i,j}\| = \|M_{t/a_i}^{[2]}F_{1,j} - F_{1,j}\| \leq \left(\frac{17}{12} + \frac{27}{16e}\right) \frac{a_i^2}{t^2}. \quad (89)$$

Thus, using (86) and (89), we have

$$\begin{aligned} \|M_t^{[2]}F - F\| &\leq \sum_{i=1}^n \sum_{j=1}^{\infty} p_{ij} \|M_t^{[2]}F_{a_i,j} - F_{a_i,j}\| \\ &\leq \left(\frac{17}{12} + \frac{27}{16e}\right) \frac{\sum_{i=1}^n (\sum_{j=1}^{\infty} p_{ij}) a_i^2}{t^2}. \end{aligned} \quad (90)$$

This completes the proof of Proposition 5.2. \square

As a consequence of the previous result, we can provide error bounds for compound distributions (that is, distribution functions of random sums, as in (14)) when the summands are mixtures of Erlang distributions, as stated in the following result.

Corollary 5.1. *Let G be the distribution function of a random sum, as in (14), in which the sequence of $(X_i)_{i \in \mathbb{N}^*}$ has a common distribution $\mathcal{ME}(a_1, \dots, a_n)$, $0 < a_1 < \dots < a_n$, as defined in (83). Let $M_t^{[2]}$ be as in (8). We have that*

$$\|M_t^{[2]}G - G\| \leq \left(\frac{17}{12} + \frac{27}{16e}\right) \frac{(1 - G(0))a_n^2}{t^2}.$$

Proof. The proof is immediate, taking into account that a mixture of Erlang distributions $\mathcal{ME}(a_1, \dots, a_n)$, $0 < a_1 < \dots < a_n$, can be expressed as in (84) and compound distributions of these random variables are also mixtures of Erlang distributions (see [19], page 106, with a slight modification in the coefficients, as we allow a point mass at 0), that is, we can write

$$G(x) = q_0 + \sum_{j=1}^{\infty} q_j F_{a_n, j}(x), \quad x \geq 0,$$

in which $\{q_j, j = 0, 1, \dots\}$ form a probability mass function (obviously, $q_0 = G(0)$). The result follows using the above expression and Proposition 5.2. \square

The class of phase-type distributions, of great importance in applied probability, can be expressed as mixtures of Erlang distributions. A phase-type distribution is defined as the time until absorption in a continuous-time Markov chain with one absorbent state (see, for instance, [12], Chapter II or [8], Chapter VIII, and the references therein). A phase-type distribution can be expressed in terms of a matrix exponential as follows. Consider a vector $\alpha = (\alpha_1, \dots, \alpha_n)$ of non-negative numbers such that $\alpha_1 + \dots + \alpha_n \leq 1$. Let A be a $n \times n$ matrix with negative diagonal entries, non-negative off-diagonal entries and non-positive row sums. A non-negative random variable X is a phase-type distribution $PH(\alpha, A)$ if its distribution function can be written as

$$F(x) = 1 - \alpha e^{xA} \mathbf{1}', \quad x \geq 0,$$

in which $\mathbf{1}'$ represents the transpose of the n th dimensional vector $\mathbf{1} = (1, \dots, 1)$. Note that phase-type distributions are absolutely continuous random variables when $\alpha_1 + \dots + \alpha_n = 1$, having positive mass at 0 (of magnitude $1 - (\alpha_1 + \dots + \alpha_n)$) when $\alpha_1 + \dots + \alpha_n < 1$. Phase-type distributions have been extensively studied from both theoretical and practical points of view. For instance, it is well known that phase-type distributions have rational Laplace transforms, thus allowing numerical computation using our approximation procedures. Also, in the next proposition, we will give an expression of phase-type distributions in terms of mixtures of Erlang distributions. This, together with Proposition 5.2, provides our approximations with rates of convergence. The proof of the next result is based on the following property of phase-type distributions, due to Maier (see [13], page 591). Let f be the density of an absolutely continuous phase-type distribution. There exists some $c > 0$ verifying

$$c_j := \frac{d^j}{dx^j} e^{cx} f(x) \Big|_{x=0} > 0, \quad j \in \mathbb{N}. \quad (91)$$

We are now in a position to state the following.

Proposition 5.3. *Let F be a phase-type distribution $PH(\alpha, A)$, with $\alpha_1 + \dots + \alpha_n > 0$. Let $c > 0$ be such that the absolutely continuous part of F satisfies the property (91).*

Then, F can be expressed as a mixture of Erlang distributions, that is,

$$F(x) = p_0 + \sum_{j=1}^{\infty} p_j F_{c,j}(x), \quad x \geq 0, \quad (92)$$

in which $p_0 = 1 - (\alpha_1 + \cdots + \alpha_n)$.

Proof. To prove (a), assume first that F is absolutely continuous, that is, that $\alpha_1 + \cdots + \alpha_n = 1$. Its density is then given by $f(x) = -\alpha e^{xA} \mathbf{A} \mathbf{1}'$, $x > 0$. We choose a $c > 0$ verifying (91). Note that we can write

$$e^{cx} f(x) = -\alpha e^{x(cI-A)} \mathbf{A} \mathbf{1}', \quad x \geq 0. \quad (93)$$

It can be easily checked that the function $-\alpha e^{x(cI-A)} \mathbf{A} \mathbf{1}'$, $x \in \mathbb{R}$ is analytic in \mathbb{R} , so if we consider the Taylor series expansion of this function around 0 and take into account (91) and (93), we have

$$e^{cx} f(x) = \sum_{j=0}^{\infty} c_j \frac{x^j}{j!}, \quad x > 0,$$

from which we can write (recall (3))

$$f(x) = \sum_{j=0}^{\infty} \frac{c_j}{c^{j+1}} \frac{c^{j+1} x^j e^{-cx}}{j!} = \sum_{j=0}^{\infty} \frac{c_j}{c^{j+1}} f_{c,j+1}(x), \quad x > 0,$$

and, in this way, we obtain the expression of f in terms of a mixture of Erlang densities with shape parameter c (by construction, the coefficients are non-negative and integrating both sides in the above expression, we see that their sum is 1). As a consequence, we can write

$$F(x) = \sum_{j=1}^{\infty} \frac{c_{j-1}}{c^j} F_{c,j}(x), \quad x \geq 0, \quad (94)$$

thus having expressed F as a mixture of Erlang distributions, as in (92). Now, assume that $0 < \alpha_1 + \cdots + \alpha_n < 1$. This means that F has a point mass at 0 of magnitude $p_0 := 1 - (\alpha_1 + \cdots + \alpha_n)$. The absolutely continuous part of F (F^{ac}) is a phase-type distribution ($PH(\bar{\alpha}, A)$) with $\bar{\alpha} = (\alpha_1 + \cdots + \alpha_n)^{-1} \alpha$. Let $c > 0$ be such that F^{ac} verifies property (93). We can write, thanks to (94),

$$F(x) = p_0 + (1 - p_0) F^{\text{ac}}(x) = p_0 + \sum_{j=1}^{\infty} (1 - p_0) \frac{c_{j-1}}{c^j} F_{c,j}(x), \quad x \geq 0.$$

This completes the proof of Proposition 5.3. \square

Remark 5.1. Expansions similar to those given in Proposition 5.3 can be found in [12], page 58. These expansions are obtained using a representation $PH(\alpha, A)$ of the distribution under consideration. Note that if we denote by $\|A\|$ the maximum absolute value of the entries of A , then it is easy to check using (93) (see [14], page 751) that $c = \|A\|$ verifies (91). However, as the representation of a phase-type distribution is not unique, this value might not be the optimum one. Also, observe that the error bound given in (85) indicates that we should take c to be as small as possible. This problem, then, is closely connected to Conjecture 6 in [14], concerning the minimum c satisfying (91) and its relation with a phase-type representation having $\|A\|$ as small as possible. To the best of our knowledge, this conjecture remains unsolved.

Acknowledgments

I would like to thank José Garrido for suggesting the final applications to phase-type distributions when I was at Concordia University and to two anonymous referees for helpful comments. This research has been partially supported by research Grants 2006-CIE-05 (University of Zaragoza), MTM2007-63683 and PR 2007-0295 (Spanish Government), E64 (DGA) and by FEDER funds.

References

- [1] Abate, J. and Whitt, W. (1995). Numerical inversion of Laplace transforms of probability distributions. *ORSA Journal of Computing* **7** 36–43.
- [2] Abate, J. and Whitt, W. (1996). An operational calculus for probability distributions via Laplace Transforms. *Adv. in Appl. Probab.* **28** 75–113. [MR1372332](#)
- [3] Abramowitz, M. and Stegun, I.A. (1964). *Handbook of Mathematical Functions*. Washington, DC: National Bureau of Standards. [MR0167642](#)
- [4] Adell, J.A. and de la Cal, J. (1993). On the uniform convergence of normalized Poisson mixtures to their mixing distribution. *Statist. Probab. Lett.* **18** 227–232. [MR1241619](#)
- [5] Adell, J.A. and de la Cal, J. (1994). Approximating gamma distributions by normalized negative binomial distributions. *J. Appl. Probab.* **31** 391–400. [MR1274795](#)
- [6] Adell, J.A. and Sangüesa, C. (1999). Direct and converse inequalities for positive linear operators on the positive semi-axis. *J. Austral. Math. Soc. Ser. A* **66** 90–103. [MR1658683](#)
- [7] Alzer, H. (1997). On some inequalities for the gamma and psi functions. *Math. Comp.* **66** 373–389. [MR1388887](#)
- [8] Asmussen, S. (2000). *Ruin Probabilities*. Singapore: World Scientific. [MR1794582](#)
- [9] Embrechts, P., Grübel, R. and Pitts, S.M. (1993). Some applications of the fast Fourier transform algorithm in insurance mathematics. *Statist. Neerlandica* **47** 59–75. [MR1208036](#)
- [10] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications II*, 2nd ed. New York: Wiley.
- [11] Grübel, R. and Hermesmeier, R. (2000). Computation of compound distributions II: Discretization errors and Richardson extrapolation. *Astin Bull.* **30** 309–331. [MR1946567](#)
- [12] Latouche, G. and Ramaswami, V. (1999). *Introduction to Matrix Analytic Methods in Stochastic Modelling*. Philadelphia: ASA-SIAM. [MR1674122](#)

- [13] Maier, R.S. (1991). The algebraic construction of phase-type distributions. *Comm. Statist. Stochastic Models* **7** 573–602. [MR1139070](#)
- [14] O’Cinneide, C.A. (1999). Phase-type distributions: Open problems and a few properties. *Comm. Statist. Stochastic Models* **15** 731–757. [MR1708454](#)
- [15] Qi, F., Cui, R., Chen, C. and Guo, B. (2005). Some completely monotonic functions involving polygamma functions and an application. *J. Math. Anal. Appl.* **310** 303–308. [MR2160691](#)
- [16] Sangüesa, C. (2008). Error bounds in approximations of compound distributions using gamma-type operators. *Insurance Math. Econom.* **42** 484–491. [MR2404310](#)
- [17] Sangüesa, C. (2008). Uniform error bounds in continuous approximations of non-negative random variables using Laplace Transforms. Preprint. Available at: <http://www.unizar.es/galdeano/preprints/2008/prep08-01.pdf>.
- [18] Sundt, B. (2002). Recursive evaluation of aggregate claims distributions. *Insurance Math. Econom.* **30** 297–322. [MR1921109](#)
- [19] Willmot, G.E. and Woo, J.K. (2007). On the class of Erlang mixtures with risk theoretical applications. *N. Am. Actuar. J.* **11** 99–105. [MR2380721](#)

Received November 2008